

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Masayuki NAKAMURA et al.
International Application No.: PCT/JP04/007307
International Filing Date: May 21, 2004
For: WEB-ENABLED ELECTRONICS APPARATUS, WEB
PAGE PROCESSING METHOD, AND PROGRAM

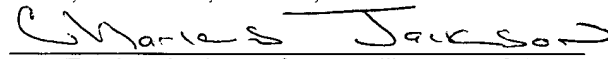
745 Fifth Avenue
New York, NY 10151

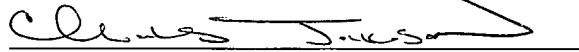
EXPRESS MAIL

Mailing Label Number: EV375019321US

Date of Deposit: January 26, 2005

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" Service under 37 CFR 1.10 on the date indicated above and is addressed to Mail Stop PCT, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.


(Typed or printed name of person mailing paper or fee)


(Signature of person mailing paper or fee)

CLAIM OF PRIORITY UNDER 37 C.F.R. § 1.78(a)(2)


Mail Stop PCT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Pursuant to 35 U.S.C. 119, this application is entitled to a claim of priority to Japan
Application No. 2003-149958 filed 27 May 2003.

Respectfully submitted,

FROMMER LAWRENCE & HAUG LLP
Attorneys for Applicants

By: 
William S. Frommer
Reg. No. 25,506
Tel. (212) 588-0800

21.05.2004

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: 2003年 5月27日

出願番号
Application Number: 特願2003-149958
[ST. 10/C]: [JP 2003-149958]

出願人
Applicant(s): ソニー株式会社

REC'D 08 JUL 2004

WIPO

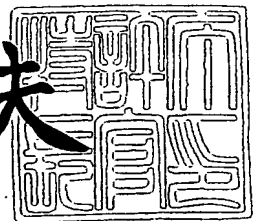
PCT

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2004年 6月21日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



Best Available Copy

出証番号 出証特2004-3054528

【書類名】 特許願

【整理番号】 0390115303

【あて先】 特許庁長官 殿

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 中村 雅之

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 由谷 哲夫

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 永留 和敏

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 山中 信太郎

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代理人】

【識別番号】 100104215

【弁理士】

【氏名又は名称】 大森 純一

【選任した代理人】

【識別番号】 100104411

【弁理士】

【氏名又は名称】 矢口 太郎

【手数料の表示】

【予納台帳番号】 069085

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0008872

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 ウェブ対応電子機器装置、ウェブページ処理方法およびプログラム

【特許請求の範囲】

【請求項 1】 少なくとも見出しとこの見出しに関連する本文を含む第 1 のウェブページを取得するウェブページ取得手段と、

前記ウェブページ取得手段により取得された前記第 1 のウェブページから前記本文を抽出して、この本文を含む第 2 のウェブページを作成するとともに、前記第 1 のウェブページから前記見出しを抽出し、この見出しを含み、かつ前記第 2 のウェブページへのリンクが設けられた第 3 のウェブページを作成するウェブページ再構築手段と

を具備することを特徴とするウェブ対応電子機器装置。

【請求項 2】 請求項 1 に記載のウェブ対応電子機器装置であって、
前記ウェブページ再構築手段は、

前記第 1 のウェブページを内部的に描画し、この描画データに基づき前記第 1 のウェブページにおける個々の表示要素の位置を判定する表示要素位置判定手段と、

前記判定された表示要素の位置に基づいて、前記個々の表示要素をレイアウト的に繋ぎの強いものどうしを結び付け、いくつかのクラスタに分別するクラスタ分別手段と、

前記個々のクラスタのレイアウト上の特徴を検出し、この特徴検出結果に基づいて前記第 1 のウェブページにおける前記見出しおよび前記本文の各クラスタを他のクラスタに対して判別する特定クラスタ判別手段と、

前記判別された前記見出しおよび前記本文の各クラスタについて、表示要素である文字の属性が同一であるクラスタどうしを集めてグループ化し、それぞれのグループに含まれるクラスタ内の文字数の平均を算出し、平均値の高いグループを前記本文として決定し、前記平均値の低いグループを前記見出しとして決定する手段と

を具備することを特徴とするウェブ対応電子機器装置。

【請求項3】 請求項2に記載のウェブ対応電子機器装置であって、
前記特定クラスタ判別手段は、

最も多くの前記表示要素と交わるページ上の縦線を重心線として決定し、この決定された重心線を基準に、前記個々のクラスタのレイアウト上の特徴を、少なくとも、左寄り、右寄り、その中間のいずれかから判定し、特徴が中間として判定されたクラスタを前記見出しおよび前記本文の各クラスタとして他のクラスタに対して判別することを特徴とするウェブ対応電子機器装置。

【請求項4】 請求項1に記載のウェブ対応電子機器装置であって、
前記本文は、前記見出しに属する記事へのリンクの一覧を含むことを特徴とするウェブ対応電子機器装置。

【請求項5】 処理演算部と、ウェブページを表示する表示部とを有するウェブ対応電子機器装置のウェブページ処理方法であって、

少なくとも見出しとこの見出しに関連する本文を含む第1のウェブページをネットワークを通じて取得するステップと、

前記処理演算部による処理演算によって、前記取得した前記第1のウェブページから前記本文を抽出し、この本文を含む第2のウェブページを作成するステップと、

前記処理演算部による処理演算によって、前記第1のウェブページから前記見出しを抽出し、この見出しを含み、かつ前記第2のウェブページへのリンクが設けられた第3のウェブページを作成するステップと

を具備することを特徴とするウェブ対応電子機器装置のウェブページ処理方法

。

【請求項6】 請求項5に記載のウェブ対応電子機器装置のウェブページ処理方法であって、

前記処理演算部は、前記第1のウェブページを内部的に描画し、この描画データに基づき前記第1のウェブページにおける個々の表示要素の位置を判定し、この判定した表示要素の位置に基づいて、前記個々の表示要素をレイアウト的に繋ぎの強いものどうしを結び付け、いくつかのクラスタに分別し、前記個々のクラスタのレイアウト上の特徴を検出し、この特徴検出結果に基づいて前記第1の

ウェブページにおける前記見出しおよび前記本文の各クラスを他のクラスに対して判別し、この判別した見出しおよび本文の各クラスについて、表示要素である文字の属性が同一であるクラスどうしを集めてグループ化し、それぞれのグループに含まれるクラス内の文字数の平均を算出し、平均値の高いグループを前記本文として決定し、前記平均値の低いグループを前記見出しとして決定することを特徴とするウェブ対応電子機器装置のウェブページ処理方法。

【請求項 7】 請求項 6 に記載のウェブ対応電子機器装置のウェブページ処理方法であって、

前記処理演算部は、最も多くの前記表示要素と交わるページ上の縦線を重心線として決定し、この決定された重心線を基準に、前記個々のクラスのレイアウト上の特徴を、少なくとも、左寄り、右寄り、その中間のいずれかから判定し、特徴が中間として判定されたクラスを前記見出しおよび前記本文の各クラスとして他のクラスに対して判別することを特徴とするウェブ対応電子機器装置のウェブページ処理方法。

【請求項 8】 請求項 5 に記載のウェブ対応電子機器装置のウェブページ処理方法であって、

前記本文は、前記見出しに属する記事へのリンクの一覧を含むことを特徴とするウェブ対応電子機器装置のウェブページ処理方法。

【請求項 9】 少なくとも見出しとこの見出しに関連する本文を含む第 1 のウェブページを取得するウェブページ取得手段と、

前記ウェブページ取得手段により取得された前記第 1 のウェブページから前記本文を抽出して、この本文を含む第 2 のウェブページを作成するとともに、前記第 1 のウェブページから前記見出しを抽出し、この見出しを含み、かつ前記第 2 のウェブページへのリンクが設けられた第 3 のウェブページを作成するウェブページ再構築手段としてコンピュータを機能させることを特徴とするプログラム。

【請求項 10】 請求項 9 に記載のプログラムであって、

前記ウェブページ再構築手段は、

前記第 1 のウェブページを内部的に描画し、この描画データに基づき前記第 1 のウェブページにおける個々の表示要素の位置を判定する表示要素位置判定手段

と、

前記判定された表示要素の位置に基づいて、前記個々の表示要素をレイアウト的に繋がり強いものどうしを結び付け、いくつかのクラスタに分別するクラスタ分別手段と、

前記個々のクラスタのレイアウト上の特徴を検出し、この特徴検出結果に基づいて前記第1のウェブページにおける前記見出しおよび前記本文の各クラスタを他のクラスタに対して判別する特定クラスタ判別手段と、

前記判別された前記見出しおよび前記本文の各クラスタについて、表示要素である文字の属性が同一であるクラスタどうしを集めてグループ化し、それぞれのグループに含まれるクラスタ内の文字数の平均を算出し、平均値の高いグループを前記本文として決定し、前記平均値の低いグループを前記見出しとして決定する手段としてコンピュータを機能させることを特徴とするプログラム。

【請求項11】 請求項10に記載のプログラムであって、

前記特定クラスタ判別手段は、

最も多くの前記表示要素と交わるページ上の縦線を重心線として決定し、この決定された重心線を基準に、前記個々のクラスタのレイアウト上の特徴を、少なくとも、左寄り、右寄り、その中間のいずれかから判定し、特徴が中間として判定されたクラスタを前記見出しおよび前記本文の各クラスタとして他のクラスタに対して判別する手段としてコンピュータを機能させることを特徴とするプログラム。

【請求項12】 請求項9に記載のプログラムであって、

前記本文は、前記見出しに属する記事へのリンクの一覧を含むことを特徴とするプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、ネットワークへの接続機能をもつPDA、携帯電話、テレビジョンセットなどの電子機器装置に適用され、ウェブ上のコンテンツを処理して、自身の表示環境に合わせて最適に表示を行うウェブ対応電子機器装置、ウェブページ

処理方法、およびプログラムに関する。

【0002】

【従来の技術】

一般に、PDA(Personal Digital(Data) Assistants)、携帯電話、テレビジョンセットといったプログラム組み込み型のウェブ対応電子機器装置が持つマン・マシンインタフェースは、パーソナルコンピュータのそれに比べると貧弱であると言わざるを得ない。これに対し、Web上のコンテンツは、マウスや高解像度の表示装置を用いたパーソナルコンピュータによる閲覧表示を想定してデザインされる場合が多く、このため上記のようなウェブ対応電子機器装置でWeb上のコンテンツを閲覧表示しようとする、ユーザは様々な不具合に遭遇することを避けて通れなかった。

【0003】

たとえば、ウェブ対応電子機器装置の表示デバイスは、パーソナルコンピュータのそれよりも解像度が低いものが採用される場合がほとんどである。前述したように、ウェブページの多くはパーソナルコンピュータに用いられる高解像度の表示デバイスでの閲覧表示を想定してデザインされているため、たとえば、図16に示すように、PDAのようなウェブ対応電子機器装置161に備えられた低解像度の表示デバイス162では、全体のウェブページ163の一部164しか一度に表示できない場合が多く、ページ全体を見るには縦横のスクロール操作を繰り返さなければならないなど、ユーザにとって操作の負担が大きかった。

【0004】

小さい画面に表示できる情報量を高める方法には、ウェブブラウザによって、画像の省略、文字詰め、文字の折り返しなどを行う方法や、表示画面の面積に応じて、表示される文字のフォントの最適なサイズを選択する技術などがある(たとえば、特許文献1参照。)

【0005】

【特許文献1】

特開2002-156957号公報(段落[0065]、図15)。

【0006】

【発明が解決しようとする課題】

しかし、ウェブブラウザによって、画像の省略、文字詰め、文字の折り返しなどを行っても、PDAなどが持つ小さい画面では、やはりページ全体の一部しか一度に表示できない。また、表示画面の面積に応じて、表示される文字のフォントの最適なサイズを選択する技術によっても、1画面に表示できる文字数にはやはり限界があり、また、文字が小さくなると読み辛くなるという弊害が発生する。

【0007】

本発明は、このような課題を解決するためになされるもので、ネットワークを通じて取得したウェブページを、低解像度の表示環境での閲覧に適したウェブページに再構築して表示させることができるとともに、少ないコストで、様々な種類の言語で記述されたウェブページの再構築を行うことのできるウェブ対応電子機器装置、ウェブページ処理方法およびプログラムを提供することを目的とする。

【0008】**【課題を解決するための手段】**

上記目的を達成するために、本発明のウェブ対応電子機器装置は、少なくとも見出しとこの見出しに関連する本文を含む第1のウェブページを取得するウェブページ取得手段と、前記ウェブページ取得手段により取得された前記第1のウェブページから前記本文を抽出して、この本文を含む第2のウェブページを作成するとともに、前記第1のウェブページから前記見出しを抽出し、この見出しを含み、かつ前記第2のウェブページへのリンクが設けられた第3のウェブページを作成するウェブページ再構築手段とを具備するものである。

【0009】

すなわち、このウェブ対応電子機器装置では、ネットワークを通じて取得した、見出しとこの見出しに関連する本文を含む第1のウェブページを、本文へのリンクが設けられた見出しのウェブページ（第3のウェブページ）と、本文のウェブページ（第2のウェブページ）とに分けて別画面で閲覧することを可能とした。これにより、PDAなどの携帯端末における貧弱な（低解像度の）表示環境に

において、パーソナルコンピュータ向けにデザインされた高解像度のウェブページ全体の内容を、スクロール操作無し、あるいは少ないスクロール量で効率良く閲覧することが可能になった。

【0010】

また、第1のウェブページの見出しがヘッドラインと小見出しとで構成され、本文が、ヘッドラインの本文と小見出しに属する記事へのリンクの一覧で構成される場合には、第3のウェブページとして、ヘッドラインの本文のページと小見出しに属する記事へのリンクの一覧ページとを作成し、第2のウェブページとして、本文へのリンクが設けられたヘッドラインと、リンク一覧ページへのリンクが設けられた小見出しとを含むページを作成するようにする。これにより、第3のウェブページでヘッドラインが指定されたなら当該ヘッドラインの本文ページを表示し、小見出しが指定されたなら当該小見出しに属する記事へのリンクの一覧ページを表示できるようになる。このように各ウェブページが全体として一定の規則性をもった形で提供されるので、ユーザにとっては、目的のウェブページに辿り着くための操作上の試行錯誤を排除することができ、ウェブページの内容そのものに集中したウェブ閲覧が可能になる。

【0011】

また、本発明のウェブ対応電子機器装置において、ウェブページ再構築手段は、第1のウェブページを内部的に描画し、この描画データに基づき第1のウェブページにおける個々の表示要素の位置を判定する表示要素位置判定手段と、判定された表示要素の位置に基づいて、個々の表示要素をレイアウト的に繋がり強いものどうしを結び付け、いくつかのクラスタに分別するクラスタ分別手段と、個々のクラスタのレイアウト上の特徴を検出し、この特徴検出結果に基づいて第1のウェブページにおける見出しおよび本文の各クラスタを他のクラスタに対して判別する特定クラスタ判別手段と、判別された見出しおよび本文の各クラスタについて、表示要素である文字の属性が同一であるクラスタどうしを集めてグループ化し、それぞれのグループに含まれるクラスタ内の文字数の平均を算出し、平均値の高いグループを本文として決定し、平均値の低いグループを見出しとして決定する手段とを具備するものである。

【0012】

ウェブページの記述言語には、HTML、XHTML、XML+CSSなど様々な種類があるが、この発明によれば、解釈して描画できるページ記述言語であればウェブページの再構築を行うことができ、タグの意味的な繋がり解析によってウェブページを再構築する方法に比べ、ページ再構築に要するコストを抑えることができる。

【0013】

さらに、本発明のウェブ対応電子機器装置において、特定クラスタ判別手段は、最も多くの表示要素と交わるページ上の縦線を重心線として決定し、この決定された重心線を基準に、個々のクラスタのレイアウト上の特徴を、少なくとも、左寄り、右寄り、その中間のいずれかから判定し、特徴が中間として判定されたクラスタを見出しおよび本文の各クラスタとして他のクラスタに対して判別する。

【0014】

ほとんどのウェブページにおいて、主要なコンテンツはページの横軸上の中間に配置されている。最も多くの表示要素と交わるページ上の縦線は、前記の主要なコンテンツが配置されたページ上の横軸上の位置と考えることができ、この縦線を重心線として、この重心線を基準に個々のクラスタのレイアウト上の特徴を、少なくとも、左寄り、右寄り、その中間のいずれかから判定すれば、特徴が中間として判定されたクラスタを見出しおよび本文の各クラスタとして他のクラスタに対して、高い精度で判別することが可能である。

【0015】

また、本発明の別の観点に基づくウェブページ処理方法は、処理演算部と、ウェブページを表示する表示部とを有するウェブ対応電子機器装置のウェブページ処理方法であって、少なくとも見出しとこの見出しに関連する本文を含む第1のウェブページをネットワークを通じて取得するステップと、処理演算部による処理演算によって、取得した第1のウェブページから本文を抽出し、この本文を含む第2のウェブページを作成するステップと、処理演算部による処理演算によって、第1のウェブページから前記見出しを抽出し、この見出しを含み、かつ第2

のウェブページへのリンクが設けられた第3のウェブページを作成するステップとを具備するものである。

【0016】

すなわち、この発明のウェブページ処理方法では、ネットワークを通じて取得した、見出しとこの見出しに関連する本文を含む第1のウェブページを、本文へのリンクが設けられた見出しのウェブページ（第3のウェブページ）と、本文のウェブページ（第2のウェブページ）とに分けて別画面で閲覧することを可能とした。これにより、PDAなどの携帯端末における貧弱な（低解像度の）表示環境において、パーソナルコンピュータ向けにデザインされた高解像度のウェブページ全体の内容を、スクロール操作無し、あるいは少ないスクロール量で効率良く閲覧することが可能になった。

【0017】

また、この発明のウェブページ処理方法において、処理演算部は、第1のウェブページを内部的に描画し、この描画データに基づき第1のウェブページにおける個々の表示要素の位置を判定し、この判定した表示要素の位置に基づいて、個々の表示要素をレイアウト的に繋がり強いものどうしを結び付け、いくつかのクラスタに分別し、個々のクラスタのレイアウト上の特徴を検出し、この特徴検出結果に基づいて第1のウェブページにおける見出しおよび本文の各クラスタを他のクラスタに対して判別し、この判別した見出しおよび本文の各クラスタについて、表示要素である文字の属性が同一であるクラスタどうしを集めてグループ化し、それぞれのグループに含まれるクラスタ内の文字数の平均を算出し、平均値の高いグループを前記本文として決定し、前記平均値の低いグループを前記見出しとして決定するようにした。

【0018】

したがって、この発明によれば、解釈して描画できるページ記述言語であればウェブページの再構築を行うことができ、タグの意味的な繋がり解析によってウェブページを再構築する方法に比べ、ページ再構築に要するコストを抑えることができる。

【0019】

さらに、この発明のウェブページ処理方法において、処理演算部は、最も多くの表示要素と交わるページ上の縦線を重心線として決定し、この決定された重心線を基準に、個々のクラスタのレイアウト上の特徴を、少なくとも、左寄り、右寄り、その中間のいずれかから判定し、特徴が中間として判定されたクラスタを見出しおよび本文の各クラスタとして他のクラスタに対して判別するようにした。

【0020】

最も多くの表示要素と交わるページ上の縦線は、前記の主要なコンテンツが配置されたページ上の横軸上の位置と考えることができ、この縦線を重心線として、この重心線を基準に個々のクラスタのレイアウト上の特徴を、少なくとも、左寄り、右寄り、その中間のいずれかから判定すれば、特徴が中間として判定されたクラスタを見出しおよび本文の各クラスタとして他のクラスタに対して、高い精度で判別することが可能となる。

【0021】

さらに、本発明の別の観点に基づくプログラムは、少なくとも見出しとこの見出しに関連する本文を含む第1のウェブページを取得するウェブページ取得手段と、ウェブページ取得手段により取得された第1のウェブページから本文を抽出して、この本文を含む第2のウェブページを作成するとともに、第1のウェブページから見出しを抽出し、この見出しを含み、かつ第2のウェブページへのリンクが設けられた第3のウェブページを作成するウェブページ再構築手段としてコンピュータを機能させるものである。

【0022】

この発明のプログラムによれば、ネットワークを通じて取得した、見出しとこの見出しに関連する本文を含む第1のウェブページを、本文へのリンクが設けられた見出しのウェブページ（第3のウェブページ）と、本文のウェブページ（第2のウェブページ）とに分けて別画面で閲覧することを可能とした。これにより、PDAなどの携帯端末における貧弱な（低解像度の）表示環境において、パーソナルコンピュータ向けにデザインされた高解像度のウェブページ全体の内容を、スクロール操作無し、あるいは少ないスクロール量で効率良く閲覧することが

可能になった。

【0023】

また、この発明のプログラムにおいて、ウェブページ再構築手段は、第1のウェブページを内部的に描画し、この描画データに基づき前記第1のウェブページにおける個々の表示要素の位置を判定する表示要素位置判定手段と、判定された表示要素の位置に基づいて、個々の表示要素をレイアウト的に繋ぎの強いものどうしを結び付け、いくつかのクラスタに分別するクラスタ分別手段と、個々のクラスタのレイアウト上の特徴を検出し、この特徴検出結果に基づいて第1のウェブページにおける見出しおよび本文の各クラスタを他のクラスタに対して判別する特定クラスタ判別手段と、判別された見出しおよび本文の各クラスタについて、表示要素である文字の属性が同一であるクラスタどうしを集めてグループ化し、それぞれのグループに含まれるクラスタ内の文字数の平均を算出し、平均値の高いグループを前記本文として決定し、平均値の低いグループを見出しとして決定する手段としてコンピュータを機能させるものである。

【0024】

この発明によれば、解釈して描画できるページ記述言語であればウェブページの再構築を行うことができ、タグの意味的な繋がり解析によってウェブページを再構築する方法に比べ、ページ再構築に要するコストを抑えることができる。

【0025】

さらに、この発明のプログラムにおいて、特定クラスタ判別手段は、最も多くの前記表示要素と交わるページ上の縦線を重心線として決定し、この決定された重心線を基準に、前記個々のクラスタのレイアウト上の特徴を、少なくとも、左寄り、右寄り、その中間のいずれかから判定し、特徴が中間として判定されたクラスタを前記見出しおよび前記本文の各クラスタとして他のクラスタに対して判別する手段としてコンピュータを機能させることを特徴とする。

【0026】

この発明によれば、特徴が中間として判定されたクラスタを見出しおよび本文の各クラスタとして高い精度で判別することが可能となる。

【0027】

【発明の実施の形態】

以下、本発明の実施形態を図面に基づき説明する。

【0028】

図1は本発明の一実施形態にかかるウェブ対応電子機器装置の電氣的な構成を示すブロック図である。

【0029】

同図に示すように、このウェブ対応電子機器装置100は、処理演算部としてのCPU (Central Processing Unit) 1、メインメモリ2、プログラム／データ記憶部3、インターネットなどのネットワーク4との接続を処理するネットワークインタフェース部5、ユーザに情報を視覚的に提供する表示デバイス6、表示デバイス6の画面への描画処理をVRAM (video RAM) 7を用いて行うグラフィックコントローラ8、ジョグダイヤルなどのユーザの操作入力部9からの入力を処理するユーザインタフェースコントローラ10、そして以上の各部の間で信号を伝達するためのバス11を備える。

【0030】

CPU 1は、たとえばプログラム／データ記憶部3に記憶されたプログラムやデータ、ユーザによる操作入力部9からの入力などに基づき、メインメモリ2をワークエリアとして用いて各種の演算処理や制御を実行する。メインメモリ2は、たとえば、RAM (Random Access Memory) などの、ランダムに読み取りと書き込みが可能な高速なメモリからなる。プログラム／データ記憶部3は、読み取り専用あるいは読み書き可能な不揮発性の記憶装置であり、たとえばROM (Read Only Memory)、フラッシュROM、ディスクドライブなどである。

【0031】

表示デバイス6は、具体的には、CRT (Cathode Ray Tube)、LCD (Liquid Crystal Display)、PDP (Plasma Display Panel)、OEL (Organic Electroluminescence) などである。ユーザの操作入力部9は、具体的には、簡易キーボード、IR (Infrared) リモートコントローラ、ジョグダイヤル、プッシュボタンなどである。

【0032】

ネットワークインタフェース部 5 は、たとえばアナログモデム、LAN (Local Area Network)、ISDN (Integrated Services Digital Network)、ADSL (Asymmetric Digital Subscriber Line)、FTTH (Fiber-To-The-Home)、Bluetooth、FOMA (W-CDMA) などである。

【0033】

このウェブ対応電子機器装置 100 では、取得したウェブページを自身の表示装置の解像度などの表示環境に適した形態に再構築して、表示閲覧する機能を備えている。

【0034】

プログラム／データ記憶部 3 には、このウェブ対応電子機器装置 100 を動作させるための OS (Operating System) などの基本プログラムのほか、この基本プログラムの下でウェブページの再構築を実行するページ再構築プログラム、ウェブブラウザなどが格納されており、これらのプログラムは、メインメモリ 2 にロードされて CPU 1 によって解釈、実行される。

【0035】

図 2 は、当該ページ再構築プログラムのモジュール構成を示す図である。同図に示すように、ページ再構築プログラム 21 は、アダプテーションプロキシ 31、アダプテーションエンジン 32、およびクラスタリングエンジン 33 で構成されている。

【0036】

次に、このページ再構築プログラム 21 の処理手順を説明する。図 3 は、このページ再構築プログラム 21 の処理手順を示すフローチャートである。なお、ここではウェブ対応電子機器装置 100 の一例として、PDA などの携帯端末 100 を考える。

【0037】

まず、この携帯端末 100 において、URL がユーザより入力される。この際、図 4 に示すように、URL (例: <http://www.somewhere.com>) 51 の後ろにページ再構築のための識別子 (例: `"/??ID=index"`) 52 を付加し、ページ閲覧要求を入力することで、この要求はページ再構築プログラムにページ再構築の要求と

して与えられる (ST301)。

【0038】

ページ再構築プログラム21は、ページ再構築の要求を受けると、アダプテーションプロキシ31を起動させ、これにURLを渡す。アダプテーションプロキシ31は、URLに基づいてインターネットからオリジナルのウェブページ34をダウンロードし、アダプテーションエンジン32に渡す (ST302)。

【0039】

アダプテーションエンジン32は、取得したウェブページ34のソースをDOM(Document Object Model)ツリー35の形でメインメモリ2に保存し、ウェブページを内部的に描画する (表示はしない)。続いて、当該ウェブページにおける文字列や画像などの表示要素の画面上の描画位置を求め、その位置情報をタグと組み合わせてメインメモリにタグ/位置情報36として保存する (ST303)。なお、表示要素の描画位置は文字フォントのサイズ、文字数、画像サイズによって変わってくるので、文字の場合には文字フォントのサイズと文字数などを加味して描画位置が求められ、画像の場合には画像サイズなどを加味して描画位置が求められる。

【0040】

DOMツリーとは、ページ全体のタグ、文字、画像などの要素をツリー階層化して、アプリケーションによるページの検索、編集などを可能としたものである。また、DOMは、XML(eXtensible Markup Language)文書をツリー構造のノード・オブジェクトの集合としてアクセスするためのAPI(Application Programming Interface)である。XMLのAPIとしては、DOMのほかにSAX(Simple API for XML)がある。

【0041】

この後、アダプテーションエンジン32は、タグ/位置情報36をクラスタリングエンジン33に渡し、クラスタ化を行うようにクラスタリングエンジン33に指示する。クラスタリングエンジン33は、タグ/位置情報36に基づき、視覚的に繋がりの強い (距離の近い) タグ (表示要素) どうしを結び付けて、ウェブページ中の各タグ (表示要素) をいくつかの塊 (クラスタ) に分別し (ST3

04)、その分別したクラスタの情報をメインメモリ2にクラスタリスト37として保存する。

【0042】

図5の符合70は、オリジナルのウェブページ60に対して、各表示要素61a～61lのクラスタ化を行った結果を示している。71a～71lは個々のクラスタを示している。71bはウェブページ上のヘッドライン（見出し）61bのクラスタ、71cはヘッドラインの本文61cのクラスタ、71f、71h、71jはそれぞれ小見出し61f、61h、61jのクラスタ、71g、71i、71kはそれぞれ小見出しに属する記事リスト61g、61i、61kの部分のクラスタである。ヘッドライン61bおよび小見出し61f、61h、61jは他の表示要素と視覚的な繋がりを持たないため、それぞれが個々のクラスタ71b、71f、71h、71jとして生成されている。また、記事リスト61g、61i、61kは一つの小見出しに属する集合ごとに一つのクラスタ71g、71i、71kとして生成されている。これ以外にも、いくつかの表示情報がクラスタ71a、71d、71lとして得られている。

【0043】

クラスタ化の手法には、たとえば、2Dデータマイニング分野で広く知られているグリッドベースの手法を利用できる（参考：<http://www.cs.ualberta.ca/~zaiane/courses/cmput695-00/papers/wave.pdf>）。

【0044】

続いて、クラスタリングエンジン33は、生成した個々のクラスタ71a～71lについてレイアウト上の特徴を抽出して、レイアウト上の意味付けを行う。すなわち、図6に示すように、個々のクラスタ71a～71lを「左寄り」（L）、「右寄り」（R）、「ヘッドライン（小見出しも含む）」（H）、「本体（記事リンクを含む）」（B）、「未使用」（U）の5種類の意味に分類し（ST305）、その結果をアダプテーションエンジン32に渡す。このクラスタの分類については後で詳細を説明する。

【0045】

図2に戻って、アダプテーションエンジン32は、各クラスタの分類結果を基

にウェブページの再構築を行い（ST306）、再構築したページの情報38をメインメモリ2に格納する。この後、ウェブブラウザは、メインメモリ2に格納された再構築ページの情報38を読み込んで表示デバイス6の画面に表示する（ST307）。

【0046】

次に、クラスタの分類方法の詳細を説明する。

【0047】

図7はクラスタ分類の処理手順を示すフローチャート、図8は図7のクラスタ分類の処理手順において重心線を決定するための処理手順、図9は重心線決定の具体例、図10は図7のクラスタ分類の処理手順においてクラスタのレイアウト上の意味を「左寄り」「右寄り」「未使用」の中から決定する処理手順、図11は図10のクラスタのレイアウト上の意味を決定する処理の具体例、図12は図7のクラスタ分類の処理手順においてクラスタのレイアウト上の意味を「ヘッドライン（小見出しも含む）」「本体（記事リンクを含む）」のうちから決定する処理手順である。

【0048】

まず、図7のST701で、クラスタリングエンジン33は、各クラスタのレイアウト上の特徴を検出するための基準の一つとなる、ページ画面の重心線を決定する。重心線とは、ページ画面上で最も多くの表示要素が配置された、Y軸に沿った線である。

【0049】

この重心線を決定する具体的な方法としては、たとえば、図8の処理手順および図9の具体例に示すように、まず、クラスタ化されたページ画面全体81をたとえば4×4の16の領域に等分割するグリッド線82を設定する（ST801）。ページ中心部分のたとえば2×2の4つの領域について、左右（X軸方向の）いずれか一方の端から一定間隔（ Δd ）を置いたY軸方向のラインごとに、存在する表示要素の数（P）をカウントし（ST802）、最大のカウント値 P_{max} が得られたY軸方向のラインを重心線83として決定する（ST803-806）。

【0050】

このようにして重心線 83 が決定した後、図 7 の ST702 で、個々のクラスタのレイアウト上の意味を「左寄り」「右寄り」「未使用」の中から決定する処理が行われる。この処理の具体的な方法としては、たとえば、図 10 の処理手順と図 11 の具体例に示すように、まず、たとえばグリッド線 82 によって分割された 4×4 の 16 の領域の上部 4×3 の 12 個の領域において、重心線 83 と交わるクラスタ 71a、71b、71c、71f、71g、71h、71i、71j、71k のうち、最も左寄りに突出している（最小の X 座標である）クラスタ（この例では 71c）の左端の X 座標値をとる Y 軸方向のライン 121 を左境界線として判定するとともに、最も右寄りに突出している（最大の X 座標である）クラスタ（この例では 71i）の右端の X 座標値をとる Y 軸方向のライン 122 を右境界線として判定する（ST1001）。これにより、ページ画面全体 81 は左右境界線 121、122 を境にして 3 つの領域に区切られることになる。

【0051】

この後、クラスタリングエンジン 33 は、クラスタリスト 37 から、一つのクラスタの情報を取り出す（ST1002）。このクラスタの情報は、このクラスタを構成する表示要素に関する情報（タグ、位置情報）を含んでいる。クラスタリングエンジン 33 は、このクラスタの情報に基づいて、クラスタのレイアウト上の意味を「左寄り」「右寄り」「未使用」の中から次のようにして決定する。

まず、左境界線 121 と右境界線 122 との双方にまたがるクラスタの場合は（ST1003 の YES）、そのクラスタを「未使用」のクラスタとして分類し（ST1007）、クラスタリスト 37 から除外する（ST1008）。

【0052】

クラスタが、左境界線 121 より左側の領域に完全に包含される場合は（ST1004 の YES）、そのクラスタを「左寄り」のクラスタとして分類し（ST1009）、右境界線 122 より右側の領域に完全に包含される場合は（ST1004 の YES）、そのクラスタを「右寄り」のクラスタとして分類し（ST1009）、クラスタリスト 37 から除外する（ST1010）。

【0053】

また、クラスタが、左境界線 121 より左側の領域に完全には包含されないが (ST1004 の NO)、左境界線 121 または右境界線 122 の一方の境界線にまたがっている場合には (ST1005 の YES)、クラスタの重心線を計算し (ST1011)、重心線が左境界線 121 と右境界線 122 のどちらに近いかわって、すなわち左境界線 121 に近い場合には「左寄り」のクラスタとして、右境界線 122 に近い場合には「右寄り」のクラスタとして分類し (ST1012)、クラスタリストから除外する (ST1013)。

【0054】

以上の処理を、クラスタリストに登録されている、一つ一つのクラスタについて繰り返す (ST1006)。

【0055】

以上の処理で「左寄り」「右寄り」「未使用」のどれにも分類されなかったクラスタは「ヘッドライン (小見出しも含む)」または「本文 (記事へのリンクも含む)」のいずれかである。このクラスタ分類は、たとえば図 12 に示す手順で行われる。

【0056】

まず、クラスタリングエンジン 33 は、クラスタリストからクラスタの情報を取り出してメインメモリ 2 上で内部的に描画展開し (ST1201)、図 11 に示した左境界線 121 と右境界線 122 との間の領域をスキャンする (ST1202)。続いて、フォントのサイズ、色、書体、背景色など、共通の表示属性を有するクラスタ (以下「同属クラスタ」という。) を一つのグループとして決定する (ST1203)。

【0057】

次に、クラスタリングエンジン 33 は、決定したグループのなかで、クラスタ数の多い 2 つのグループを選択し (ST1204)、それぞれのグループについて、同属クラスタ内での文字数などの情報量の平均を算出する (ST1205)。これにより、情報量の平均値の高い (文字数の多い) グループを「本文 (記事リンクを含む)」として決定し、情報量の平均値の低い (文字数の少ない) グループを「ヘッドライン (小見出し含む)」として決定する (ST1206)。

【0058】

次に、ウェブページの再構築の詳細を説明する。

【0059】

アダプテーションエンジン32は、たとえば図13に示すように、ヘッドライン131と小見出し132とで構成されるトップページ133、小見出し132に属する記事へのリンク134を集めた記事リンク集ページ135、ヘッドライン131の本文や小見出し132に属する記事で構成される本文・記事ページ136、137などの再構築を行う。

【0060】

トップページ133において、ヘッドライン131がジョグダイヤルなどの操作でユーザによって選択された場合、そのヘッドライン131に設定されたハイパーリンクによって本文・記事ページ137の表示へと切り替わる。また、トップページ133において、任意の小見出し132がユーザによって選択された場合には、その選択された小見出し132に設定されたハイパーリンクによって、その小見出し132に属する記事リンク集ページ135が表示され、さらに、この記事リンク集ページ135において任意の記事へのリンク134がユーザによって選択されることで、そのリンク先の本文・記事ページ136が表示される。再度、他の本文・記事ページを表示したい場合は、ウェブブラウザの戻りボタンなどを用いて、トップページ133や記事リンク集ページ135に戻って、同様な操作を繰り返せばよい。

【0061】

これらのページのレイアウトは、あらかじめ、携帯端末の表示画面のサイズ、解像度などの表示環境に合わせて最適に設定されている。

【0062】

図14はウェブページの再構築（トップページ133の作成時）の処理手順を示すフローチャートである。

【0063】

まず、アダプテーションエンジン32は、クラスタの分類データをロードする（ST1401）。続いて、オリジナルのDOMツリー（図2の35）から降順

にタグを読み込み（ST1402）、クラスタの分類データに基づいて、オリジナルのDOMツリーから、ヘッドラインまたは小見出しのタグを検索し（ST1404）、再構築ページのDOMツリーに該当タグを追加する（ST1405）。オリジナルのDOMツリーに次の該当するタグがあれば（ST1406のYES）、ST502に戻ってそのタグを読み込み、当該次のタグがヘッドラインのタグではなければ（小見出しであれば）（ST1403のNO）、この小見出しのタグをオリジナルのDOMツリーから検索し（ST1404）、再構築ページのDOMツリーに該当タグを追加する（ST1405）。このようにして、オリジナルのDOMツリーから、トップページ133を構成するヘッドラインと小見出しのタグをそれぞれ検索して、再構築ページのDOMツリーに追加し、再構築ページを完成させる。

【0064】

記事リンク集ページ135および本文・記事ページ136について同様に、ST1404で、クラスタの分類データに基づいて、オリジナルのDOMツリーから、記事リンクおよび本文・記事を検索し、ST1405で、再構築ページのDOMツリーに該当タグを追加することによって作成することができる。そして、以上のように作成した各再構築ページに必要なリンクを設定することによって、図13に示したようなページの移動を実現することができる。

【0065】

このように本実施形態によれば、パーソナルコンピュータの表示環境向けにデザインされたウェブページを、PDAなどの携帯端末の表示環境に適したデザインに変換して表示させることができる。具体的には、携帯端末の表示画面に一度に表示可能なサイズ（解像度）のウェブページを再構築することで、スクロール操作無しでウェブページ全体の閲覧が可能になる。また、ヘッドラインと小見出しとでメインページを構成し、このトップページでヘッドラインが指定されたなら当該ヘッドラインの本文ページを表示し、小見出しが指定されたなら当該小見出しに属する記事へのリンクの一覧ページを表示させたりすることができる。このように各ウェブページが全体として一定の規則性をもった形で提供されるので、ユーザにとって効率的なウェブ閲覧が可能になる。さらに言うならば、目的の

ウェブページに辿り着くための操作上の試行錯誤を排除することができ、内容そのものに集中したウェブ閲覧が可能になる。

【0066】

また、本実施形態によれば、解釈して描画できるページ記述言語であれば、ウェブページの再構築を行うことができる。すなわち、HTML (HyperText Markup Language)、XHTML (eXtensible HyperText Markup Language)、XML + CSS (Cascading Style Sheets) など、ウェブページの記述言語には様々な種類があるが、本実施形態は、これら様々な記述言語で作成されたウェブページの再構築を同一のロジックで実現し得るものである。これに対して、タグの意味的な繋がり (セマンティクス) 解析によってウェブページを再構築する方法があるが、この方法ではページ記述言語の種類ごとに対応する解析プログラムが必要となり、また、多大な解析時間を要する。本実施形態は、このようなタグの意味的な繋がり解析によってウェブページを再構築する方法に比べ、ページ再構築に要するコストを大幅に低減することができる。

【0067】

また、本実施形態は、オリジナルのウェブページのタグを用いて再構築ページを作成するので、再構築したウェブページを既存のウェブブラウザをそのまま用いて閲覧できるという利点がある。また、言語の種類 (日本語、英語など) やロケールに依存することなく、ウェブページの再構築を行うことができる。

【0068】

なお、ページ再構築プログラムは、Web対応電子機器装置100に組み込んで用いられることに止まらず、パーソナルコンピュータやサーバ用途のコンピュータなどに、記憶媒体や通信媒体を通じて組み込み可能なプログラムとして提供してもよい。

【0069】

たとえば、図15に示すように、ページ再構築プログラムを構成するモジュールであるアダプテーションプロキシ31、アダプテーションエンジン32、クラスタリングエンジン33を、LAN (Local Area Network) やインターネットなどのネットワーク151上に存在するサーバ152に組み込んでおき、サーバ1

52が、PDAなどのウェブ対応電子機器装置であるクライアント153からのリクエストを受けて、クライアント153が指定するウェブページをウェブサイト154から取得し、ウェブページの再構築のための一連の処理を行って、再構築ページをクライアント153にネットワーク151を介して配信するように構成してもよい。

【0070】

また、アダプテーションプロキシ、アダプテーションエンジン、クラスタリングエンジンの各機構を複数のサーバに分散して配置し、複数のサーバが協働してウェブページの再構築に係る一連の処理を分散して行うように構成してもよい。

【0071】

なお、本発明は上述したいずれの実施形態にも限定されず、本発明の技術思想の範囲内で適宜変更して実施できる。

【0072】

【発明の効果】

以上説明したように本発明によれば、低解像度の表示環境において、パーソナルコンピュータ向けにデザインされた高解像度のウェブページ全体の内容を、スクロール操作無し、あるいは少ないスクロール量で効率良く閲覧することができるとともに、少ないコストで、様々な種類の言語で記述されたウェブページの再構築を行うことが可能になる。

【図面の簡単な説明】

【図1】 本発明の一実施形態にかかるウェブ対応電子機器装置の電氣的な構成を示すブロック図である。

【図2】 ページ再構築プログラムのモジュール構成を示す図である。

【図3】 ページ再構築プログラムの処理手順を示すフローチャートである。

【図4】 携帯端末においてページ再構築のための識別子を付加したURLの入力の様子を示す図である。

【図5】 オリジナルのウェブページの例と、このウェブページに対して各表示要素のクラスタ化を実施した結果を示す図である。

【図6】 クラスタのレイアウト上での意味分類の結果の例を示す図である。

【図 7】 クラスタ分類の処理手順を示すフローチャートである。

【図 8】 図 7 のクラスタ分類処理において重心線を決定するための処理手順を示すフローチャートである。

【図 9】 重心線の決定の具体例を示す図である。

【図 10】 図 7 のクラスタ分類処理においてクラスタのレイアウト上の意味を「左寄り」「右寄り」「未使用」の中から決定する処理手順を示すフローチャートである。

【図 11】 図 10 のクラスタのレイアウト上の意味を決定する処理の具体例を示す図である。

【図 12】 図 7 のクラスタ分類処理においてクラスタのレイアウト上の意味を「ヘッドライン（小見出しも含む）」「本体（記事リンクを含む）」のうちから決定する処理手順を示すフローチャートである。

【図 13】 再構築されたウェブページの例を示す図である。

【図 14】 ウェブページの再構築（トップページの作成時）の処理手順を示すフローチャートである。

【図 15】 ネットワーク上のサーバでウェブページの再構築を行う場合の構成を示すブロック図である。

【図 16】 一般的なウェブページを低解像度の表示デバイスに表示させた様子を示す図である。

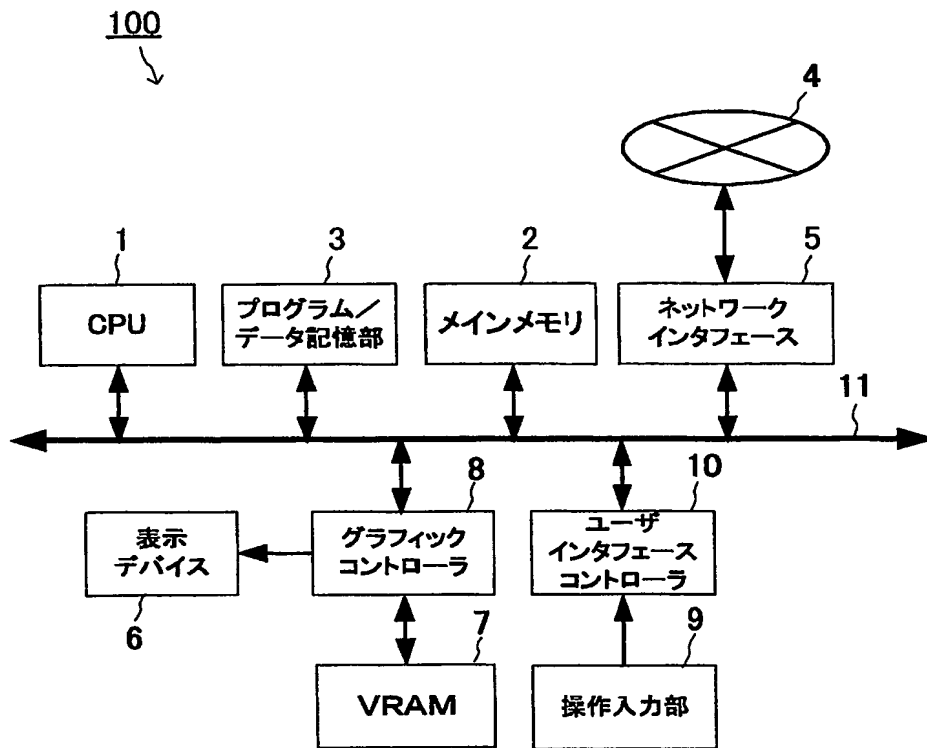
【符号の説明】

- 1 CPU
- 2 メインメモリ
- 3 プログラム／データ記憶部
- 4 ネットワーク
- 6 表示デバイス
- 21 ページ再構築プログラム
- 31 アダプテーションプロキシ
- 32 アダプテーションエンジン
- 33 クラスタリングエンジン

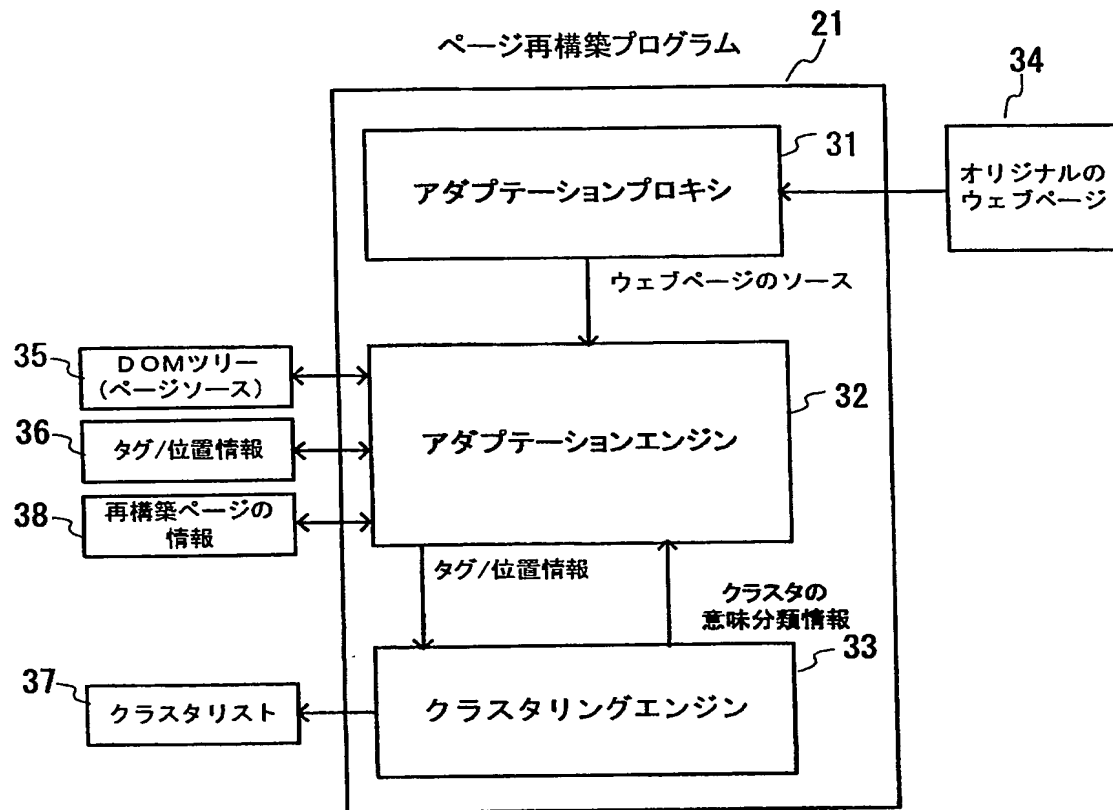
- 60 オリジナルのウェブページ
- 61a～61l オリジナルのウェブページの表示要素
- 61b オリジナルのウェブページのヘッドライン
- 61c オリジナルのウェブページの本文
- 61f、61h、61j オリジナルのウェブページの小見出し
- 61g、61i、61k オリジナルのウェブページの記事リスト
- 71a～71l クラスタ
- 100 ウェブ対応電子機器装置
- 131 再構築ウェブページのヘッドライン
- 132 再構築ウェブページの小見出し
- 133 トップページ
- 135 記事リンク集ページ
- 136 記事ページ
- 137 本文ページ

【書類名】 図面

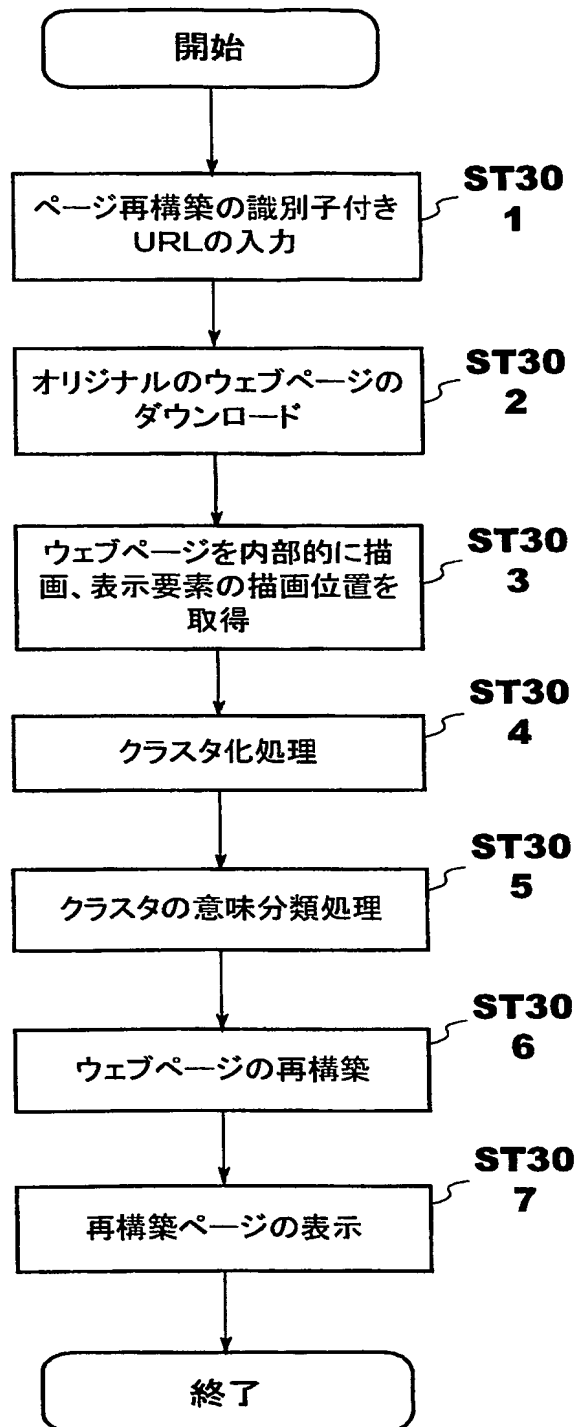
【図 1】



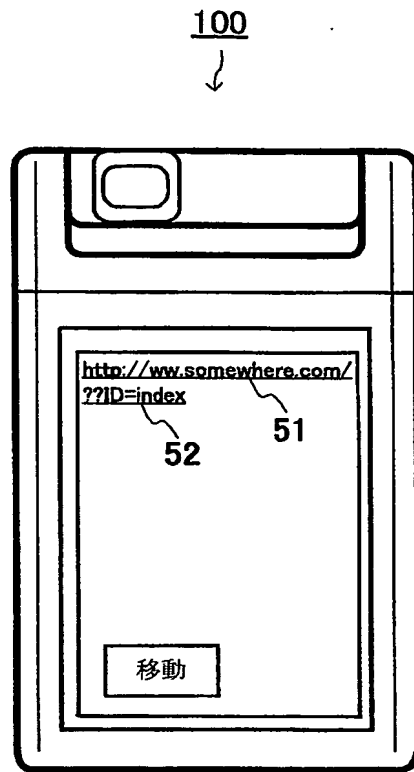
【図 2】



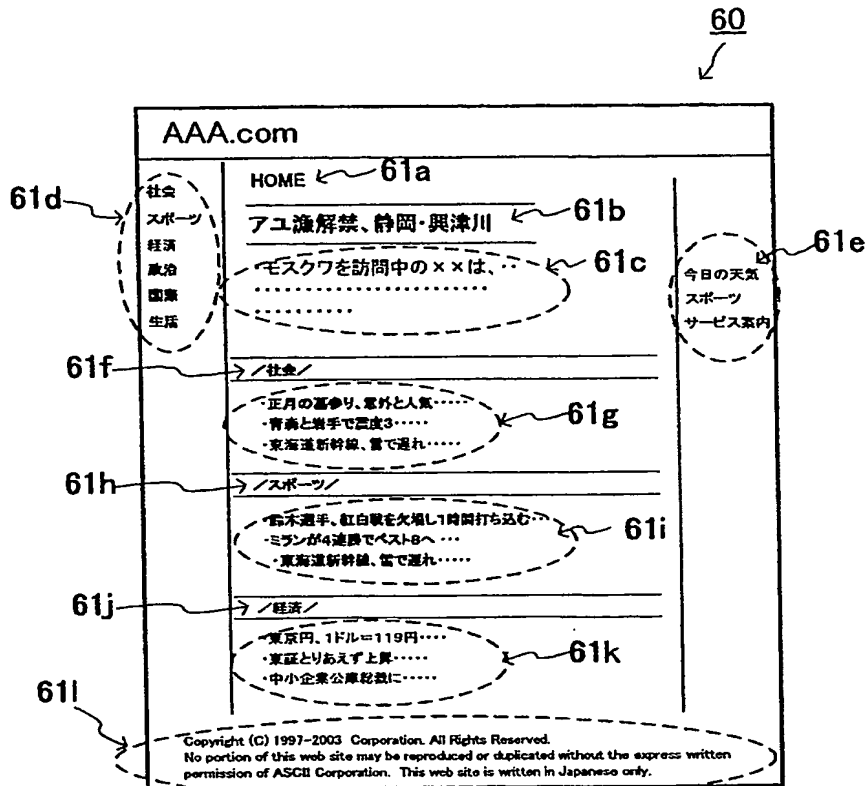
【図 3】



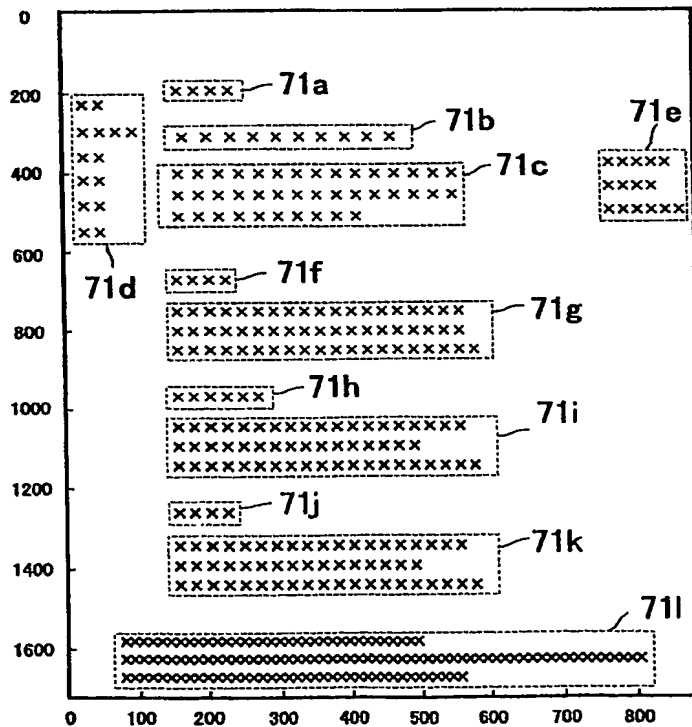
【図 4】



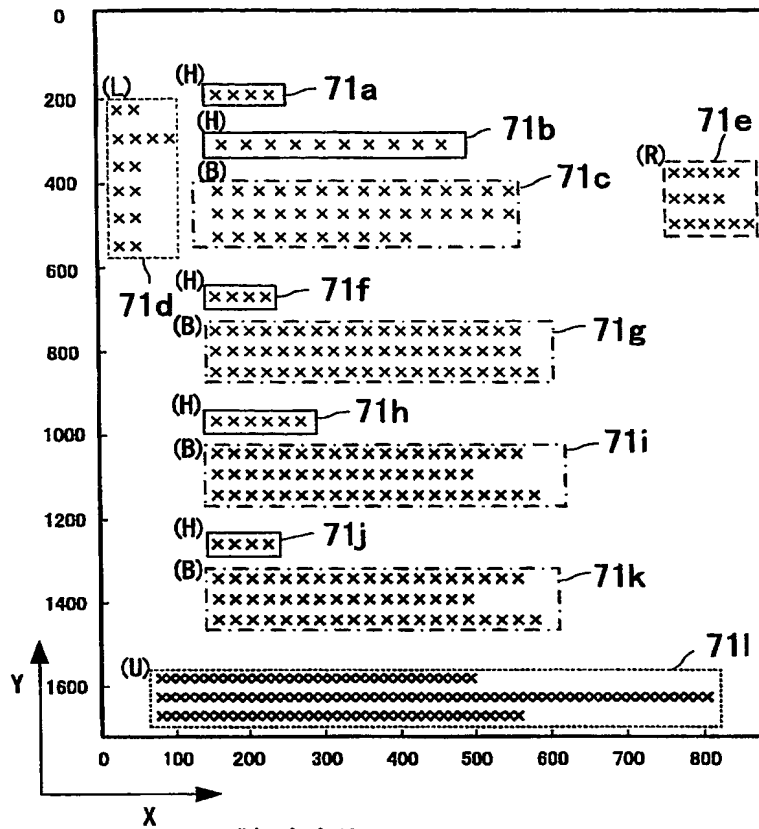
【図 5】



クラスタ化

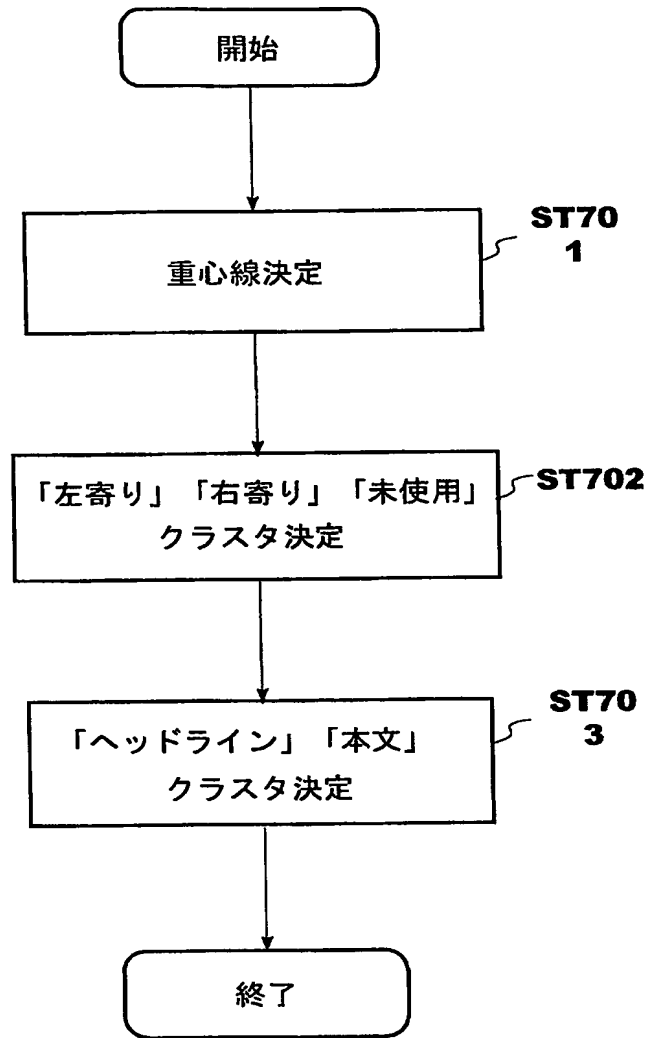


【図 6】

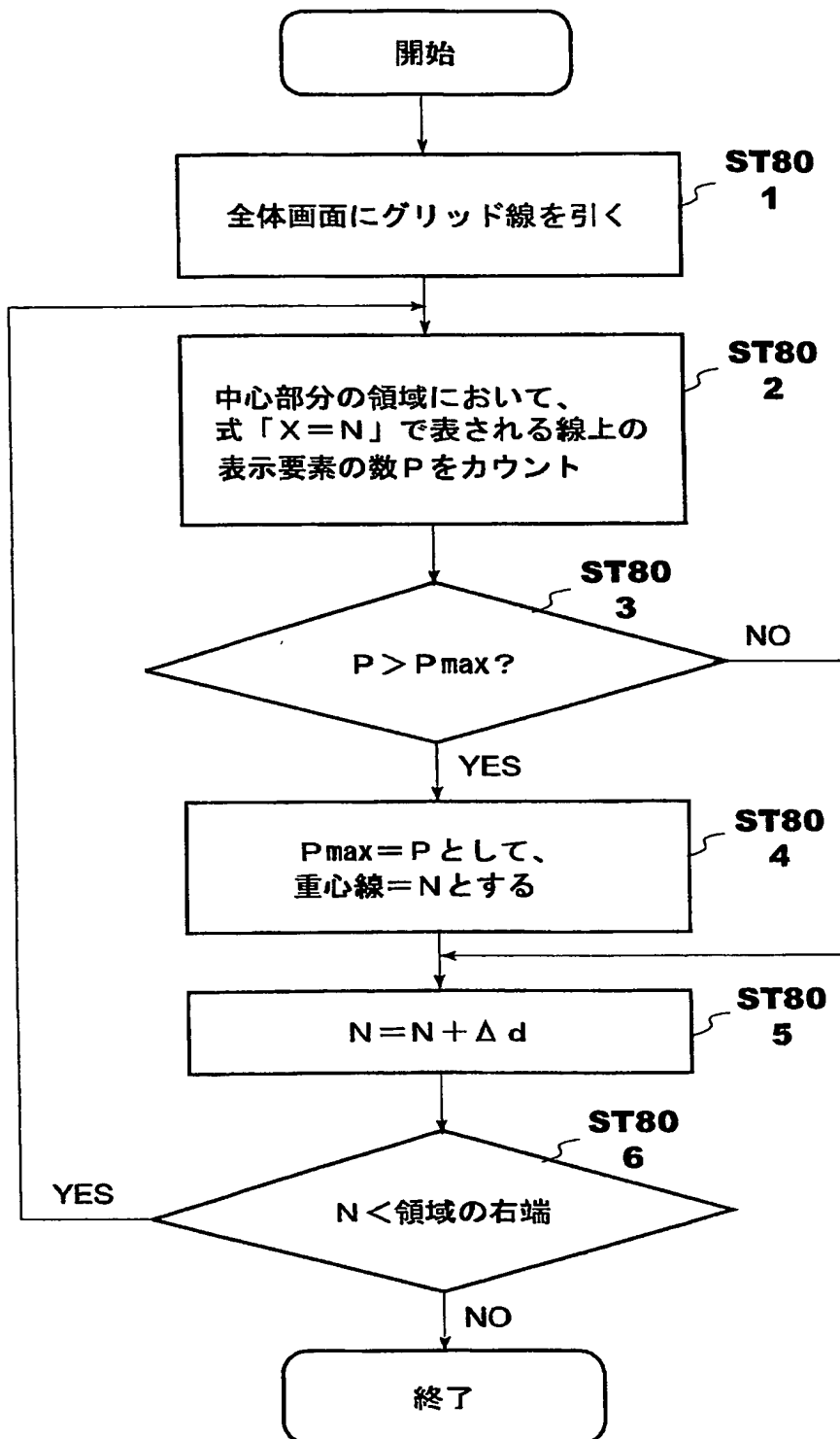


(L): 左寄り
 (R): 右寄り
 (H): ヘッドラインまたは小見出し
 (B): 本体または記事へのリンク
 (U): その他

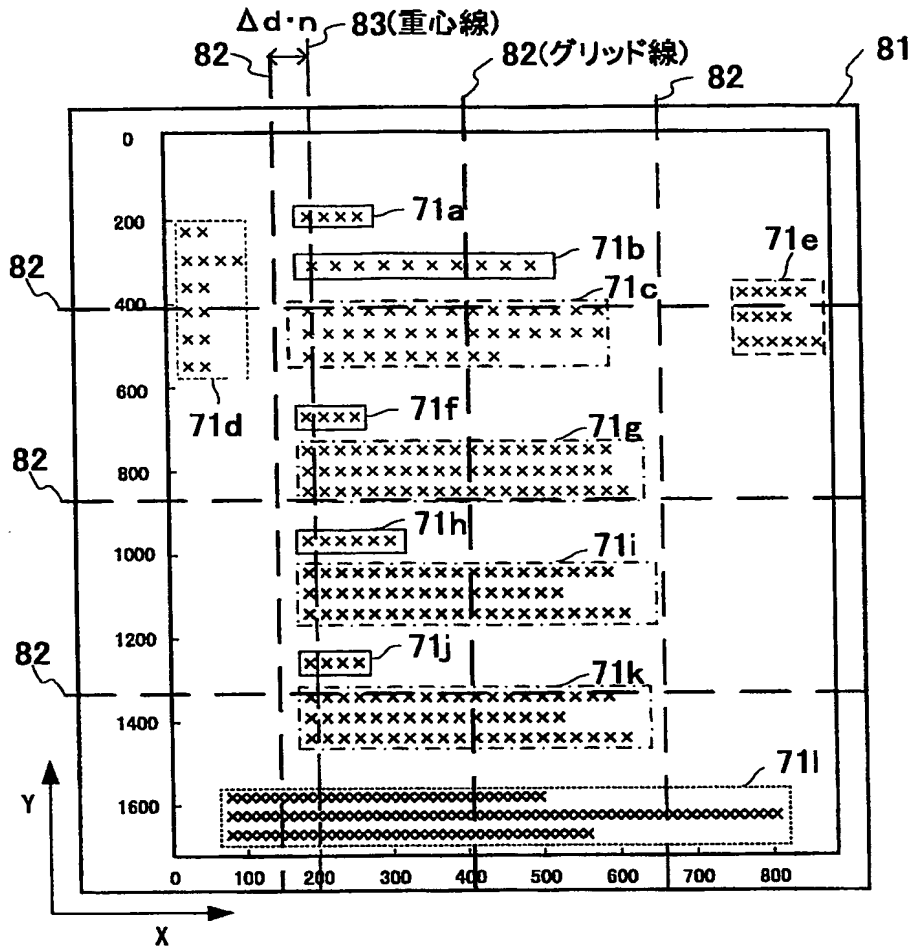
【図 7】



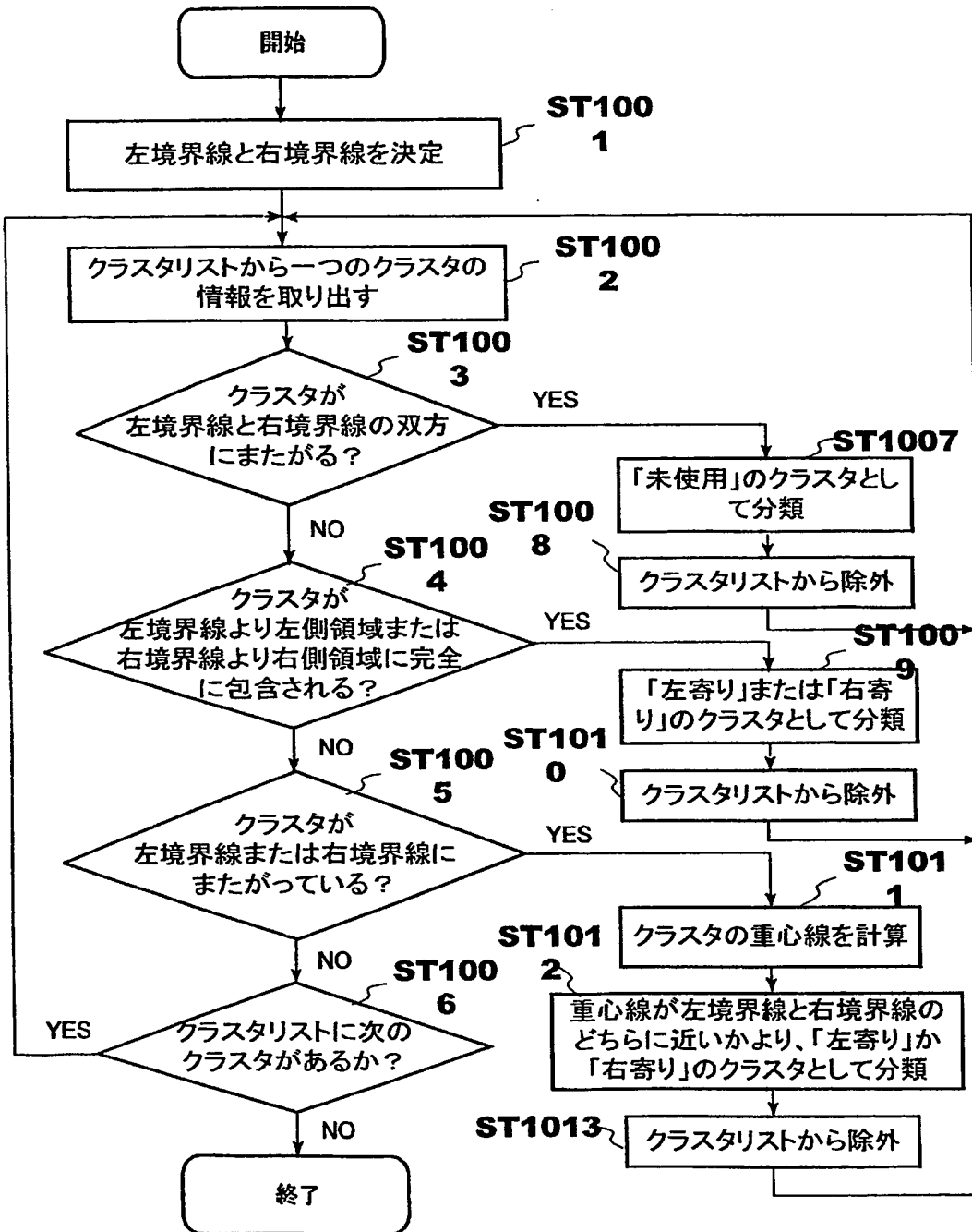
【図 8】



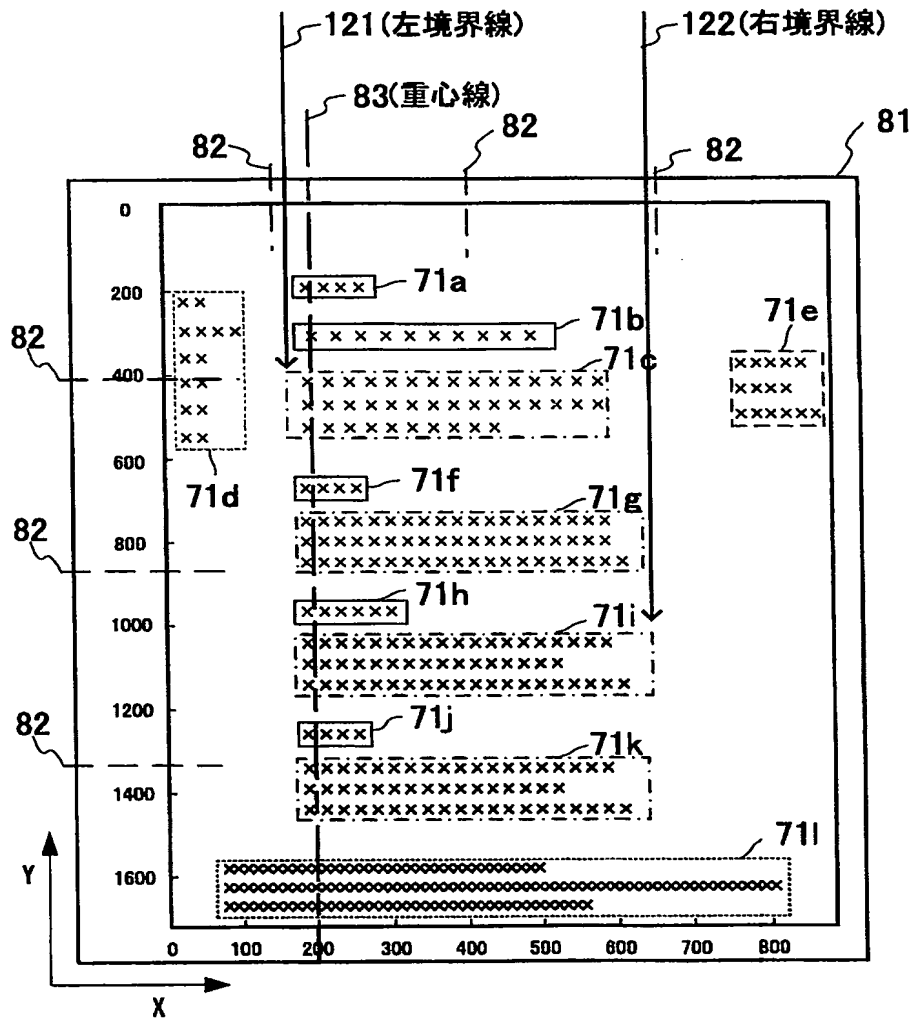
【図9】



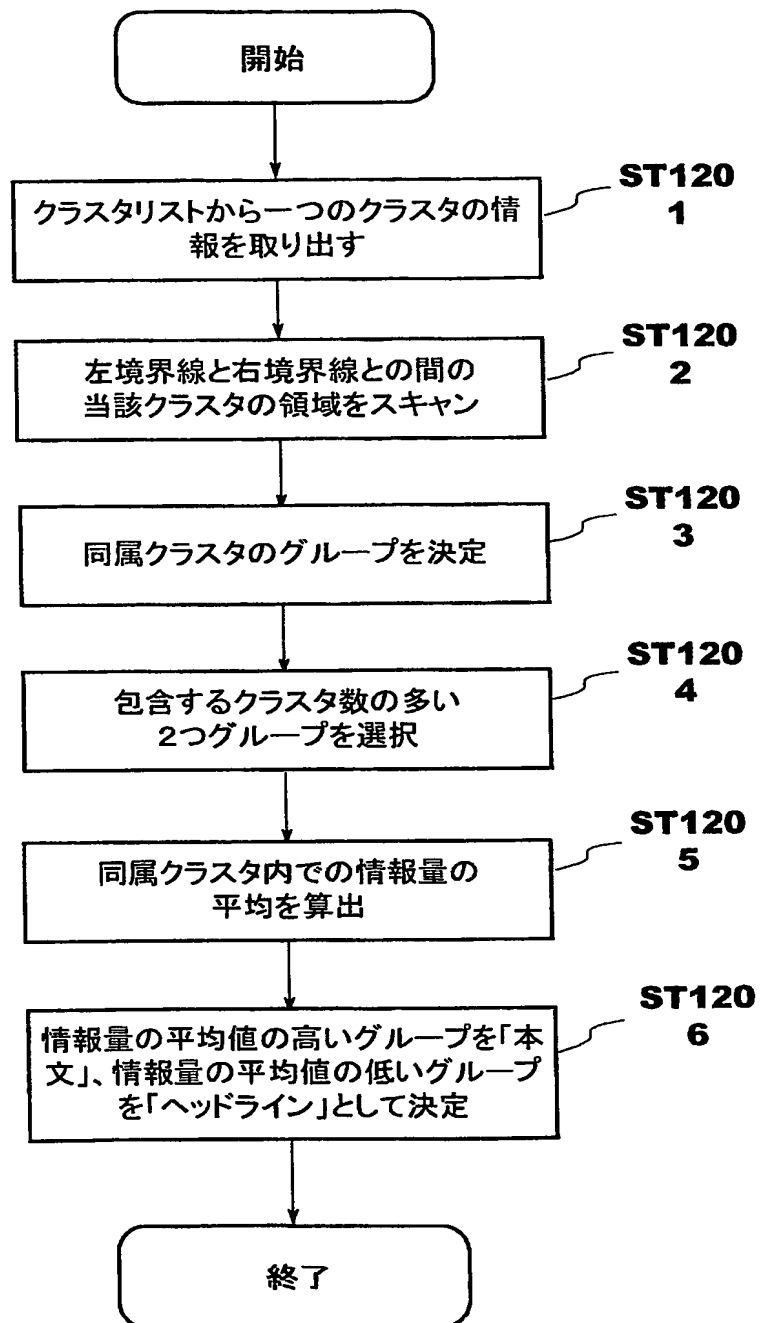
【図 10】



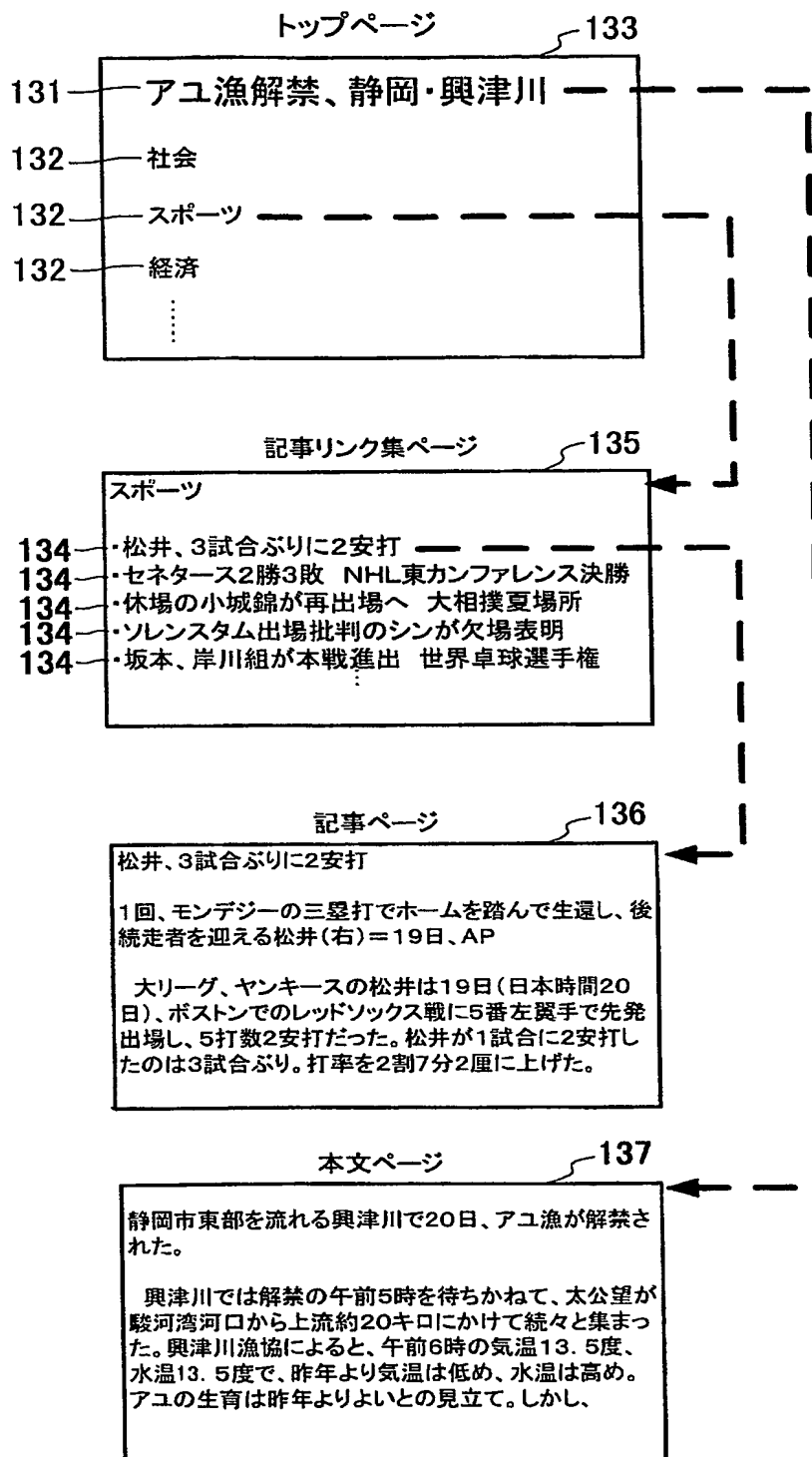
【図 11】



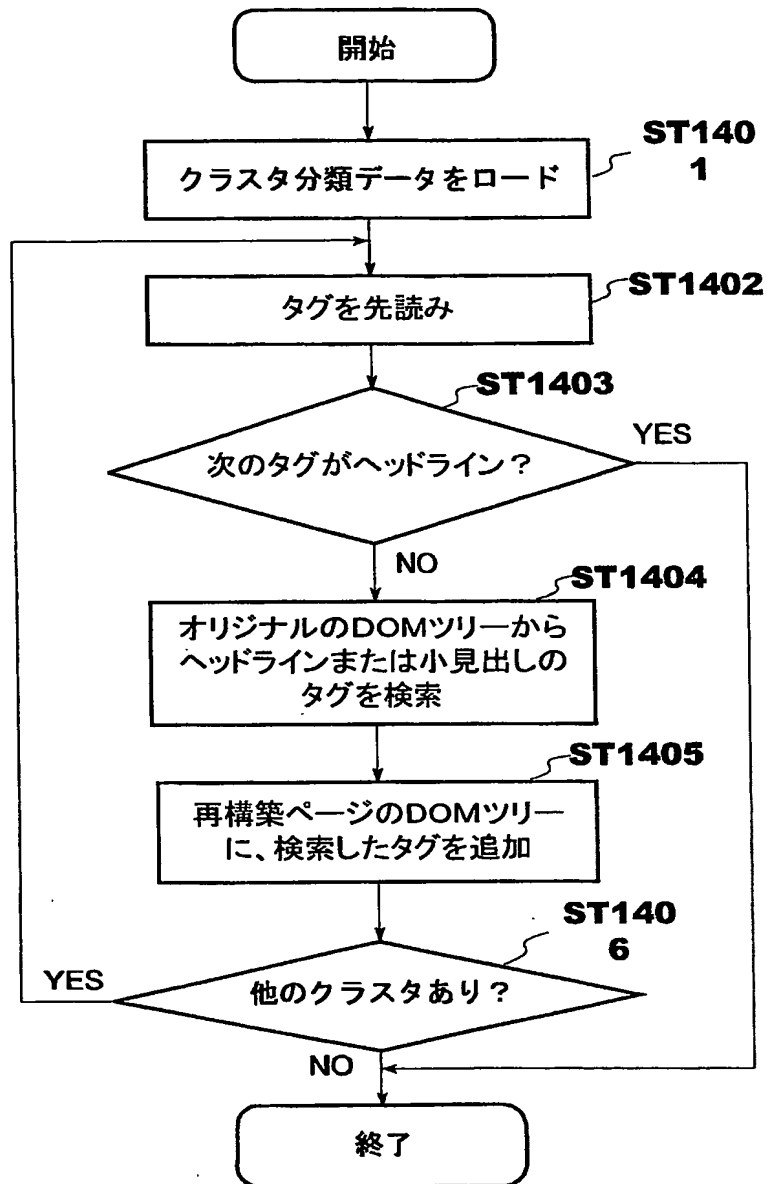
【図 12】



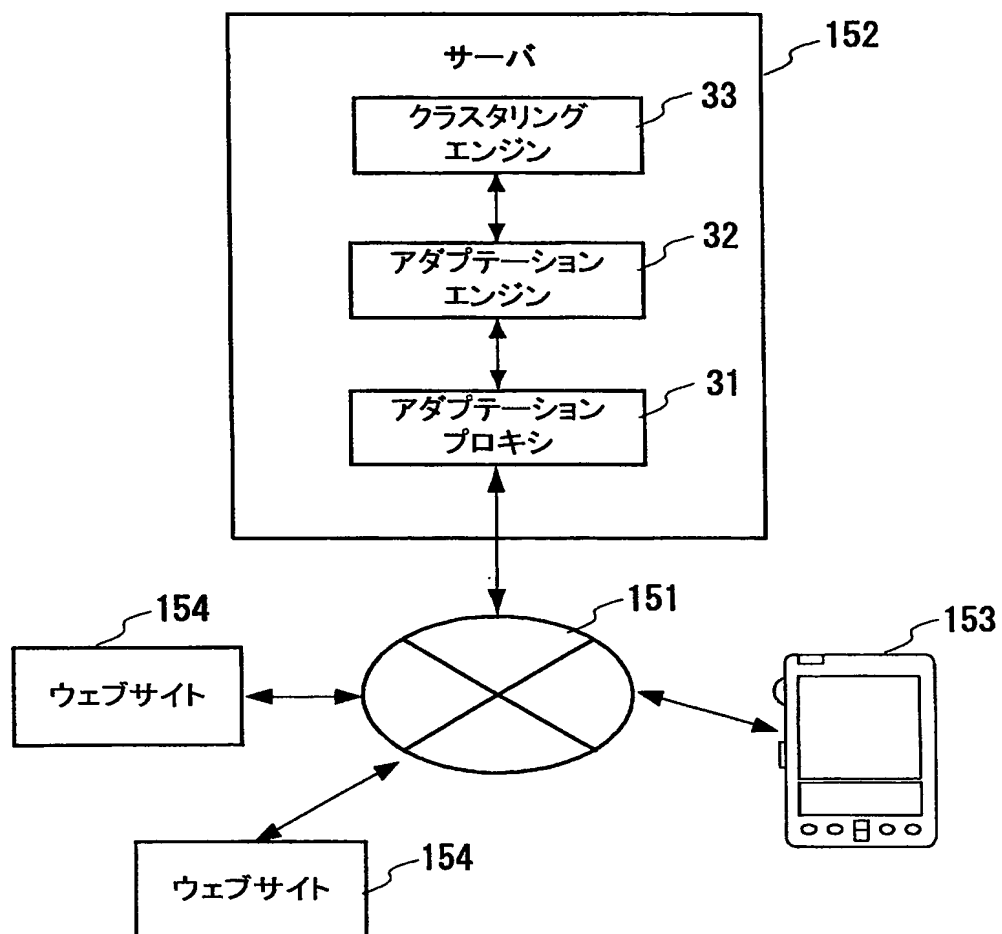
【図 13】



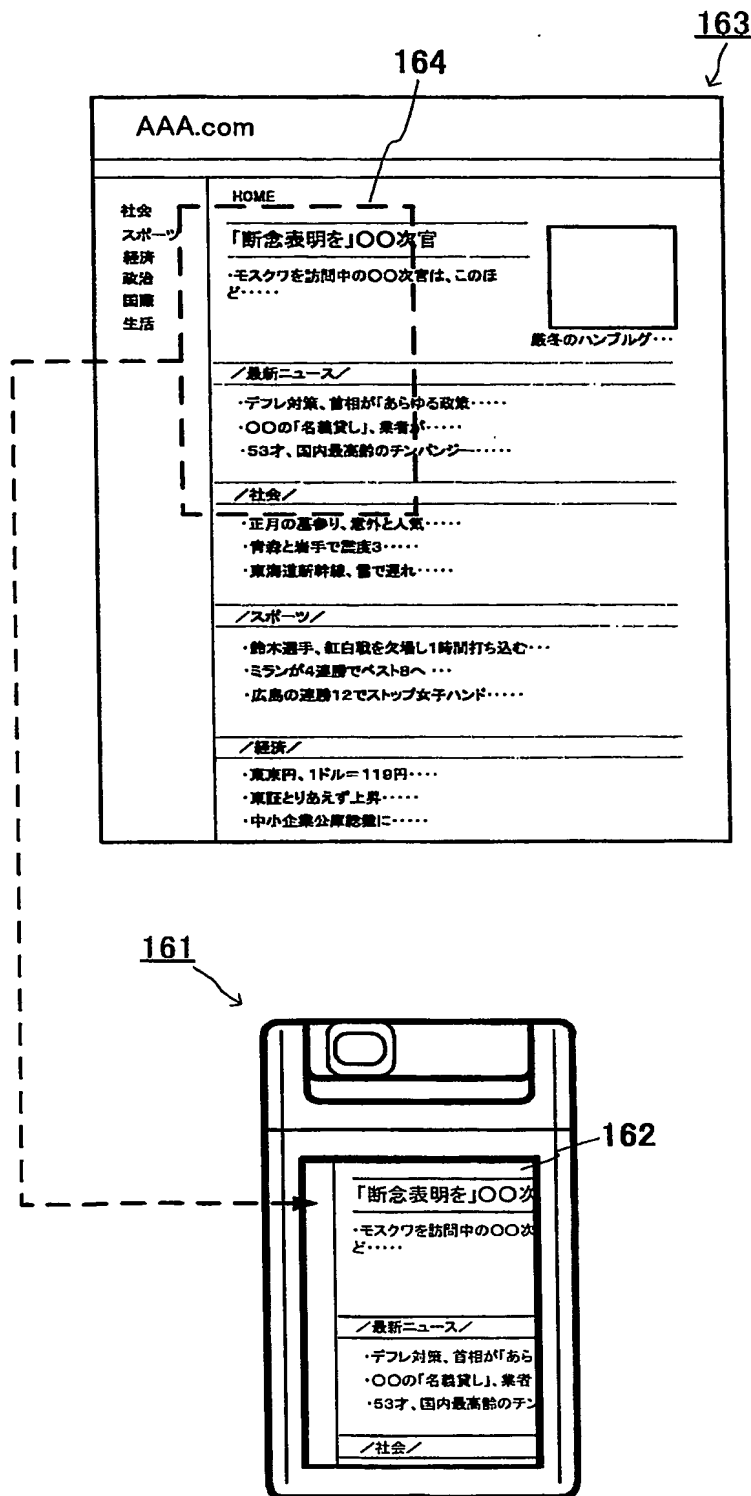
【図14】



【図 15】



【図 16】



【書類名】 要約書

【要約】

【課題】 ネットワークを通じて取得したウェブページを、低解像度の表示環境での閲覧に適したウェブページに再構築して表示させる。

【解決手段】 ヘッドラインとその本文、小見出しと小見出しに属する記事へのリンクを表示要素として含むウェブページを取得し、これを内部的に描画し、描画データに基づき各表示要素の位置を求める。求めた表示要素の位置に基づいて各表示要素をいくつかのクラスタに分別し、個々のクラスタのレイアウト上の特徴を検出し、この特徴検出結果に基づいて見出し・本文の各クラスタを判別する。次に見出し・本文の各クラスタについて文字属性が同一のクラスタどうしを集めてグループ化し、それぞれのグループに含まれるクラスタ内の文字数の平均値の高いグループを本文として、平均値の低いグループを見出しとして決定する。そして、本文・記事の各ページを作成し、これら本文・記事ページへのリンクを設けたヘッドラインおよび小見出しを含むトップページを作成する。

【選択図】 図 2

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 1 4 9 9 5 8
受付番号	5 0 3 0 0 8 7 9 6 4 9
書類名	特許願
担当官	第六担当上席 0 0 9 5
作成日	平成 1 5 年 5 月 2 8 日

< 認定情報・付加情報 >

【提出日】 平成15年 5月27日

次頁無

特願 2 0 0 3 - 1 4 9 9 5 8

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 2 1 8 5]

1. 変更年月日

1 9 9 0 年 8 月 3 0 日

[変更理由]

新規登録

住 所

東京都品川区北品川 6 丁目 7 番 3 5 号

氏 名

ソニー株式会社

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.